November, 1962

A NOTE ON ESTIMATION OF EQUATION SYSTEMS USING AN ELECTRONIC COMPUTER

J.M. Bennett, J.H. Duloy, and H.S. Konijn

Technical Report No. 17

University of Minnesota

Minneapolis, Minnesota

A NOTE ON ESTIMATION OF EQUATION SYSTEMS USING AN ELECTRONIC COMPUTER

J.M. Bennett, J.H. Duloy, and H.S. Konijn*

Summary.

For computing k-class estimates of the coefficients of a single stochastic equation in a system, a method is given which has advantages when certain electronic computers are used; this method also yields variances and covariances of the estimates and tests for the validity of the model.

1. Introduction.

Limited information estimators for the coefficients of a single stochastic equation belonging to a system of such equations were first introduced by Anderson and Rubin [1]. Their estimators were based on a suggestion by Girshick, who originally called them "reduced form" estimators. In effect they are chosen [1, equation (5.7)] so as to maximize the likelihood derived under the restriction that some of the coefficients in the system vanish (the so-called order conditions for identification), and are defined when the smallest root k_1 of the equation $|W' - k'W| = 0$ (see below) is a simple one, which will be so almost certainly under these circumstances [4, p. 173]. They are also minumum variance-ratio estimators.

By not insisting on maximizing a likelihood, these limited information estimators were generalized by Theil [6] to what he calls k-class estimators; for k=1 he calls them raw second round estimators and Basmann [2] calls them

* The first and last named authors are with the University of Sydney, the second with the University of New England in Armidale, N.S.W., Australia. From September, 1962 on, the last named author's address is: Department of Statistics, University of Minnesota.

generalized classical linear estimators. For $k=k_1$, we get the estimator mentioned in the previous paragraph. The estimators with $k=k_1$, with $k=1$, and generally with $\text{plim}\ n^{\frac{1}{2}}(k-1) = 0$, have the same asymptotic mean and variance, but since the estimators are simpler functions of the observations for k equal to a nonrandom quantity than for $k=k_1$, convergence to the common asymptotic distribution may be more rapid in the former case than in the latter. For $k=0$ we get the least squares single equation estimators, which are not usually consistent.

Reference [5, chapter X] discusses computation of k_1 estimators. In this note we discuss methods which, though inefficient when only desk calculators are used, may be considered when it is planned to use automatic computers. When such computers are available, it is often more important to economize on programming time than on running time, which, since most problems of the present type involve a small number of variables, is relatively inexpensive, even if the program is somewhat wasteful.

In particular, in electronic computation there is usually little drawback in increasing somewhat the size of a matrix to be inverted, especially if this involves scheduling fewer matrix multiplications and eliminates the need for distinguishing special cases (indirect least squares in case of exact identification, and the cases of appendices 1 and 2 of [5, chapter X]). The programming becomes quite simple if, as is usual, subroutines are available for the computation of moments, for selection of rows and columns from and bordering of matrices, and for inverting matrices. (If maximum likelihood estimates are desired, we also need a subroutine for the largest root of a symmetric positive definite matrix; in the second part of section 7 we also need the next root.)

After preparation of the final version of this report, [4] appeared which calls attention to the availability of an IBM computer program for calculation of many of the quantities discussed here (and some others).

2. Notation.

Denote the equation whose coefficients we wish to estimate by $y = Y\mathfrak{h}+P\pi+u$. Here all observed variates are measured from their mean, and

> y is an $n \times 1$ vector of observations on that endogenous variate which is normalized by assigning it in the equation to be estimated a coefficient of unity;
>
> Y is an $n \times h$ matrix of observations on the other jointly determined endogenous variates included in the equation, and $\mathfrak{h}$ an $h \times 1$ vector of coefficients to be estimated;
>
> P is an $n \times p$ matrix of predetermined variates included in this equation, and π a $p \times 1$ vector of coefficients to be estimated;
>
> u is an $n \times 1$ vector of disturbances with zero mean, variances σ^2 and zero covariances.

Define also Q, an $n \times q$ matrix of predetermined variates in the system but excluded from the equation.

For distributional assumptions see the various quoted references. Note that when maximum likelihood estimates are mentioned, they are meant to be calculated under the assumption of normality of the disturbances, but the asymptotic distributions of these estimates do not depend on normality assumptions.

3. Moments.

We first compute the matrix of sums of squares and products of all observations (if several equations are to be estimated the calculation may involve a larger set of jointly determined variates than y and Y; for subsequent estimation of a particular equation we then ignore those not appearing in that equation). Denote this matrix or n^{-1} times it -- the moment matrix -- by M_{yYPQ}, where the subscripts denote the variables included (which, as previously mentioned, are measured as deviations from their means). For the estimation of the coefficients or for tests and confidence intervals it makes no difference whether we use moments, as done below, or sums of squares and products. For the estimation of variances we have indicated the required adjustment.

From M^I_{yYPQ}, the inverse of M_{yYPQ}, partition out a (1+h) - rowed square submatrix W^I as follows:

$$M^I_{yYPQ} = \begin{array}{c} \\ 1 \\ h \\ p \\ q \end{array} \begin{array}{c} \begin{array}{cc|cc} 1 & h & p & q \end{array} \\ \left[\begin{array}{c|c} W^I & \\ \hline & \end{array}\right] \end{array}$$

In section 4b below we also make use of V^I, where V is a matrix obtained by deleting the first row and column of W (the latter is obtained as the inverse of W^I). For many computers a subroutine is available or easily written which allows efficient computation of V^I from W^I without computing W or V, neither of which is needed in what follows.

4. <u>k arbitrary</u>.

a. <u>Estimates of the coefficients</u>.

Select some k and compute

$$D_k = \begin{array}{c} \\ 1 \\ h \\ p \\ 1 \\ h \end{array} \begin{array}{c} \begin{array}{ccc|cc} 1 & h & p & 1 & h \end{array} \\ \left[\begin{array}{cc|c} \multicolumn{2}{c|}{M_{yYP}} & \begin{array}{c} k^{\frac{1}{2}}I \\ \hline 0 \end{array} \\ \hline k^{\frac{1}{2}}I & 0^T & W^I \end{array}\right] \end{array}$$

Here I is a $(1+h) \times (1+h)$ unit matrix and 0 is a $p \times (1+h)$ matrix of zeros, and 0^T the transpose of 0. We need to compute the inverse of D_k, or rather certain elements in the first column as indicated below; calling the first element α_k we have

$$D_k^I = \begin{array}{c|c|c|} & 1 & h \quad p \quad 1 \quad p \\ \hline 1 & \alpha_k & \\ h & -\alpha_k \mathfrak{h} & \\ p & -\alpha_k \tilde{\pi} & \\ 1 & & \\ h & & \\ \hline \end{array}$$

From this $\mathfrak{h}$ and $\tilde{\pi}$ are obtained by division.

b. Estimates of variances and covariances.

The matrix W^I has been obtained in section 3. W'^I is a submatrix of M_{yYP}^I:

$$M_{yYP}^I = \begin{array}{c|c|c|} & 1 \quad h & p \\ \hline 1, h & W'^I & \\ \hline p & & \\ \hline \end{array}$$

Define:

$$S_k' = \begin{array}{c|cc|c|} & 1 & h & 1 \\ \hline 1 & \multicolumn{2}{c|}{\multirow{2}{*}{W'^I}} & -1 \\ h & & & \mathfrak{h} \\ \hline 1 & -1 & \mathfrak{h}^T & 0 \\ \hline \end{array}$$

The element in the last row and column of $S_k'^I$ equals

$$- \frac{1}{\text{residual sums of squares } / \ n},$$

or, in case M represents a matrix of sums of squares and products

$$-\frac{1}{\text{residual sums of squares}},$$

and σ^2 may be estimated by

$$\tilde{\sigma}^2 = \text{residual sums of squares} \ / \ (n-1-p-q).$$

(There is no certainty about the best choice of divisor.)

To estimate the asymptotic variance - covariance matrix of $\tilde{\mathfrak{h}}$ and $\tilde{\pi}$, form an $(h+p+h) \times (h+p+h)$ submatrix, E_k of D_k, by deleting the first and the $(1+h+p+1)^{st}$ rows and the same columns of D_k, i.e., by omitting the two rows and two columns of D_k corresponding to y, and replacing the $h \times h$ submatrix in the lower right hand corner by V^I, obtained in section 3. From the inverse of E_k, partition out a submatrix C_k as follows:

$$E_k^I = \begin{array}{c} \\ h \\ p \\ h \end{array} \begin{array}{cc} h \quad p & h \\ \hline \multicolumn{1}{|c|}{C_k} & \multicolumn{1}{c|}{} \\ \hline \multicolumn{1}{|c|}{} & \multicolumn{1}{c|}{} \\ \hline \end{array}$$

For the desired variance - covariance matrix we can write $n^{-1}\sigma^2 C_k$, or, in case M represents a matrix of sums of squares and products, $\sigma^2 C_k$. Our result is asymptotic so that under the assumption made on k, $\sigma^2 C_1$ is equivalent with this. To estimate this matrix replace σ^2 by $\tilde{\sigma}^2$.

5. <u>Maximum likelihood estimation: $k=k_1$, the smallest root of $|W' - k'W| = 0$.</u>

As neither W nor W' have been obtained explicitly, and as the largest root is easier to compute than the smallest, we obtain the smallest root k_1 of $|W' - k'W| = 0$ from the largest value $1/k_1$ of $1/k'$ satisfying the equation $|k'^{-1}W^I - W'^I| = 0$. We then obtain estimates $\hat{\mathfrak{h}}$ and $\hat{\pi}$ of $\mathfrak{h}$ and π from $D_{k_1}^I$ and estimates of the variance - covariance matrix of $\hat{\mathfrak{h}}$ and $\hat{\pi}$ from $S_{k_1}'^I$ and C_{k_1}.

Note that the latter estimate is a consistent estimate for the variance - co-variance matrix of $\hat{\mathfrak{h}}$ and $\hat{\pi}$, but differs somewhat from the maximum likelihood estimate of that matrix.

Note that in the case of exact identification, $k_1=1$.

6. Least squares estimation: k=0

As is well known, in this classical case we obtain the first column of the inverse of M_{yYP} (see also the beginning of section 4b):

$$M^I_{yYP} = \begin{array}{c|c|c|}
 & 1 & h \quad p \\
\hline
1 & \alpha_o & \\
h & -\alpha_o\bar{\mathfrak{h}} & \\
p & -\alpha_o\bar{\pi} & \\
\hline
\end{array}$$

From this we obtain the least squares estimates $\bar{\mathfrak{h}}$ and $\bar{\pi}$ by division. Similarly C_o is simply M^I_{YP}.

The sum of squares of residuals is $1/\alpha_o$.

7. Tests of the coefficients of Q and of identification.

If q=h we have either exact identification or underidentification, if $q > h$ we usually have overidentification. In the latter case (assuming identification) a test has been given in [1] of the null hypothesis which asserts the correctness of the specification that the q predetermined variables with observation matrix Q have zero coefficients in the equation which is being estimated. For an improved form [3] of this test note that the distribution of

$$\varphi_k \frac{n-1-p-q}{q-h}$$

is, under the null hypothesis, asymptotically equivalent with an F variate with q-h and n-1-p-q degrees of freedom. Here φ_k+1 is the sum of squares of residuals divided by

$$[-1 \quad \mathfrak{h}^T] \quad W \quad [-1 \quad \mathfrak{h}^T]^T.$$

The ratio can be computed by dividing the element in the last row and column of the inverse of

$$S_k = \begin{array}{c|cc|c|} & 1 & h & 1 \\ \hline 1 & \multicolumn{2}{c|}{\multirow{2}{*}{W^I}} & -1 \\ \cline{4-4} h & & & \mathfrak{h} \\ \hline 1 & -1 & \mathfrak{h}^T & 0 \\ \hline \end{array}$$

by the corresponding element in the inverse of $S_k'^I$ (given in section 4b). In the case of $k=k_1$, the ratio is simply k_1.

Given that the q predetermined variables with observation matrix Q have zero coefficients in the equation being estimated and that $q \geq h$, a test has been given [5, page 184] which allows one to decide on identifiability. For a possibly improved form of this test we note that, if k_1 and k_2 are the two smallest roots of $|W' - k'W| = 0$,

$$(k_1-1)(k_2-1)\frac{n-1-p-q}{q-h+1}$$

is, under the null hypothesis of nonidentifiability, asymptotically equivalent with an F variate with q-h+1 and n-1-p-q degrees of freedom.

8. Scaling.

To prevent losses of accuracy it is advisable that units of measurement be such that all variables are of the same order of magnitude. In many computer subroutines all inputs have to be less than a certain number in absolute value. We select units of measurement so that the data satisfy both requirements. The matrices $k^{\frac{1}{2}}I$ and W^I may not satisfy the scaling condition; we may replace them in D_k by $(\ell\ k)^{\frac{1}{2}}I$ and $\ell\ W^I$ without affecting the values obtained for α_k, $-\alpha_k\mathfrak{h}$ and $-\alpha_k\pi$ or C_k.

To set up S_k or S_k', we may have to scale W^I or W'^I by multiplying by f and $[-1 \quad \mathfrak{h}^T]$ by multiplying by f', in which case the lower right hand element in the inverse is not (using moments)

$$- \frac{1}{\text{residual sum of squares } / \, n}$$

but $f(f')^{-2}$ times this.

9. Proofs.

Proofs may be supplied by the reader by using repeatedly the well-known formulae for the elements of the inverse

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

of a nonsingular partitioned matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

namely:

$$D=(d - ca^I b)^I, \quad C= - Dca^I, \quad B= - a^I bD, \quad A=a^I - a^I bC$$

for nonsingular a, and

$$A=(a - bd^I c)^I, \quad B= - Abd^I, \quad C= - d^I cB, \quad D=d^I - d^I cB$$

for nonsingular d.

Thus in section 3, if

$$M_{yYPQ} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

with $d=M_{PQ}$, nonsingular, and if W is the matrix of sums of squares and products of the residuals of the regressions of the jointly determined endogenous variables in the equation being estimated on all the predetermined variables in the system, $W=a - bd^I c$, and so $A=W^I$.

10. An Example.

Let us estimate the last equation in the system of M.A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food," *Econometrica*, vol. 15, 1947, pp. 79-110. In our relation y_5 is y, y_2 is Y, z_8 is P, and z_6, z_7 and z_9 are Q_1, Q_2 and Q_3, respectively. The matrix M_{yYPQ} of sums of squares and products of deviations from means is included in the two matrices given on page 100 and the matrix on the top of page 101, op. cit.:

	y	Y	P	Q_1	Q_2	Q_3
y	3,164.9495	1,231.2685	-306.8500	2,169.4165	6,297.5185	1,290.2350
Y	1,231.2685	583.2285	-257.7500	920.2995	1,870.0155	430.5750
P	-306.8500	-257.7500	665.0000	-415.2500	658.1500	317.7000
Q_1	2,169.4165	920.2995	-415.2500	3,071.7255	3,963.8095	1,714.9250
Q_2	6,297.5185	1,870.0155	658.1500	3,963.8095	32,367.3055	4,956.0350
Q_3	1,290.2350	430.5750	317.7000	1,714.9250	4,956.0350	2,067.0700

A submatrix of its inverse is W^I:

	y	Y
y	.003,339,800	-.006,055,432
Y	-.006,055,432	.015,125,267

and V^I is

	Y
Y	.004,146,088

The inverse of M_{yYP} has as first two columns the submatrix W'^I

	y	Y
y	.002,157,906	-.004,966,243
Y	-.004,966,243	.013,498,395

a. Least squares estimation.

Consider the first column of M^I_{yYP}. The reciprocal

$$463.412$$

of the first element is the sum of squares of residuals for least squares. The negative of this multiplied by the first and second element, respectively, gives the least squares estimates of $\mathfrak{h}$ and π:

$$\bar{\mathfrak{h}} = 2.301 \qquad \bar{\pi} = .431$$

C_o is M^I_{YP}:

	Y	P
Y	.002,069	.000,802
P	.000,802	.001,815

The residual variance is estimated by (1/17)(463.412):

$$\bar{\sigma}^2 = 27.260$$

as there are 20 observations. The estimated variance-covariance matrix of $\bar{\mathfrak{h}}$ and $\bar{\pi}$ is then

	$\bar{\mathfrak{h}}$	$\bar{\pi}$
$\bar{\mathfrak{h}}$	.0564	.0219
$\bar{\pi}$	.0219	.0495

b. <u>Minimum</u> <u>variance-ratio</u> <u>estimation</u>.

We find for the roots of $|W' - k'W| = 0$:

$$k_1 = 1.089{,}270 \qquad k_2 = 2.847{,}399$$

The first three elements in the first column of $D^I_{k_1}$ are

y	65.178
Y	-187.906
P	-42.756

so that

$$\hat{\mathfrak{h}} = 2.883 \qquad \hat{\pi} = .656$$

The lower right-hand corner element of $S'^I_{k_1}$ is

$$-.001{,}595{,}242$$

so that the estimate of residual variance is (1/15) times the negative reciprocal of this

$$\hat{\sigma}^2 = 41.791$$

C_{k_1} is

	Y	P
Y	.004,533	.001,757
P	.001,757	.002,185

and so the estimated variance-covariance matrix of $\hat{\mathfrak{h}}$ and $\hat{\pi}$ is

	$\hat{\mathfrak{h}}$	$\hat{\pi}$
$\hat{\mathfrak{h}}$	.1894	.0734
$\hat{\pi}$	.0734	.0913

c. <u>Raw second round estimation (k=1)</u>.

The first three elements in the first column of D_1^I are

y	.020,326
Y	-.056,667
P	-.012,584

so that

$$\tilde{\mathfrak{h}} = 2.788 \qquad \tilde{\pi} = .619$$

The lower right hand corner element of $S_1'^I$ is

$$-.001,730,819$$

so that the estimate of residual variance is (1/15) times the negative reciprocal of this

$$\tilde{\sigma}^2 = 38.517$$

C_1 is

	Y	P
Y	.004,130	.001,601
P	.001,601	.002,124

and so the estimated variance-covariance matrix of $\tilde{\mathfrak{h}}$ and $\tilde{\pi}$ is

	$\tilde{\mathfrak{h}}$	$\tilde{\pi}$
$\tilde{\mathfrak{h}}$	.1590	.0617
$\tilde{\pi}$	.0617	.0818

d. <u>Test of the coefficients of Q</u>.

The element in the last row and column of S_1^I is

$$-.001,891,919$$

Dividing by the corresponding element of $S_1'^I$ and subtracting 1 gives

$$\varphi_1 = .093,077$$

Multiplying this by 15/2 gives the value

$$.698$$

to the test statistic which has an asymptotic F distribution with 2 and 15 degrees of freedom when the coefficients of Q vanish. So we are not led to reject this hypothesis.

For $k = k_1$, $\varphi_{k_1} = k_1 - 1 = .089{,}270$ and the value the statistic takes on is

$$.670$$

The identifiability statistic $(k_1-1)(k_2-1)(15/3)$ takes on the value

$$.825,$$

so, given that the coefficients of Q are zero, we cannot reject the hypothesis of nonidentifiability.

REFERENCES

[1] Anderson, T.W. and H. Rubin (1949). "Estimation of the parameters of a single equation in a complete system of stochastic equations," Annals of Mathematical Statistics, 20, 46-63.

[2] Basmann, R.L. (1957). "A generalized classical method of linear estimation of coefficients in a structural equation," Econometrica, 25, 77-83.

[3] Basmann, R.L. (1960). "On finite sample distributions of generalized classical linear identifiability test statistics," Journal of the American Statistical Association, 55, 650-659.

[4] Eisenpress, H. (1962). "Note on the computation of full-information maximum-likelihood estimates of coefficients of a simultaneous system," Econometrica, 30, 343-348.

[5] Hood, W.C. and T.C. Koopmans, editors (1953). Studies in Econometric Method, Cowles Commission for Research in Economics, Monograph 14, New York and London, XIX and 333 pp.

[6] Theil, H. (1955). "The measurement of consumers' behavior," Review of the International Statistical Institute, 23, 36-45.